



PageRank in undirected random graphs

K. Avrachenkov, A. Kadavankandy,
L. Ostroumova Prokhorenkova and A. Raigorodskii

**RESEARCH
REPORT**

N° XXX

Sept 2015

Project-Team Maestro



PageRank in undirected random graphs

K. Avrachenkov, A. Kadavankandy*,
L. Ostroumova Prokhorenkova and A. Raigorodskii

Project-Team Maestro

Research Report n° XXX — Sept 2015 — 14 pages

Abstract: PageRank has numerous applications in information retrieval, reputation systems, machine learning, and graph partitioning. In this paper, we study PageRank in undirected random graphs with expansion property. The Chung-Lu random graph represents an example of such graphs. We show that in the limit, as the size of the graph goes to infinity, PageRank can be represented by a mixture of the restart distribution and the vertex degree distribution.

Key-words: PageRank, undirected random graphs, expander graphs, Chung-Lu random graphs

* Primary author, arun.kadavankandy@inria.fr

**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Le Pagerank des graphes aléatoires non-orientés

Résumé : Le PageRank éprouve de nombreuses applications sur les domaines de récupération d'informations, systèmes de réputation, apprentissage automatique, et partitionnement des graphes. Dans ce travail, nous étudions le PageRank sur les graphes aléatoires non-orientés avec une propriété de dilatation, par exemple les graphes de Chung-Lu. Nous montrons que dans la limite, lorsque la taille du graphe tend vers l'infini, le PageRank peut être représenté par un mélange de la distribution de redémarrage et la distribution de degrés des sommets du graphe.

Mots-clés : PageRank,

1 Introduction

PageRank has numerous applications in information retrieval [20, 26, 30], reputation systems [19, 21], machine learning [3, 4], and graph partitioning [1, 11]. A large complex network can often be conveniently modeled by a random graph. It is surprising that not many analytical studies are available for PageRank in random graph models. We mention the work [5] where PageRank was analysed in preferential attachment models and the more recent works [9, 10] where PageRank was analysed in directed configuration models. According to several studies [16, 18, 23, 29] PageRank and in-degree are strongly correlated in directed networks such as Web graph. Apart from some empirical studies [8, 27], to the best of our knowledge, there is no rigorous analysis of PageRank on basic undirected random graph models such as the Erdős–Rényi graph [17] or the Chung-Lu graph [13]. In this paper, we fill this gap and show that in these models PageRank can be represented as a mixture of the restart distribution and the vertex degree distribution when the size of the graph goes to infinity. First, we show the convergence in total variation norm for a general family of random graphs with expansion property. Then, we specialize the results for the Chung-Lu random graph model proving the element-wise convergence. We conclude the paper with numerical experiments and several interesting future research directions.

2 Definitions

Let $G^{(n)} = (V^{(n)}, E^{(n)})$ denote a family of random graphs, where $V^{(n)}$ is a vertex set, $|V^{(n)}| = n$, and $E^{(n)}$ is an edge set, $|E^{(n)}| = m$. Denote also by $A^{(n)}$ the associated adjacency matrix with elements

$$A_{ij}^{(n)} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise.} \end{cases}$$

In this work, we analyze PageRank on undirected graphs and hence $A^T = A$. At the same time, our analysis easily extends to some families of weighted undirected graphs. We omit the superscript index n when it is clear from the context. Let $\mathbf{1}$ be the vector of ones of an appropriate dimension and let $d = A\mathbf{1}$ be the vector of (weighted) degrees. It is helpful to define $D = \text{diag}(d)$, a diagonal matrix with the degree sequence on its diagonal. Let $P = AD^{-1}$ be the Markov transition matrix corresponding to the standard random walk on the graph and let $Q = D^{-1/2}AD^{-1/2}$ be the symmetrized transition matrix. In this paper we work with column stochastic matrices. Note that the symmetrized transition matrix is closely related to the normalized Laplacian $\mathcal{L} = I - D^{-1/2}AD^{-1/2} = I - Q$ [12]. Further we will also use the resolvent matrix $R = [I - \alpha P]^{-1}$ and the symmetrized resolvent matrix $S = [I - \alpha Q]^{-1}$.

Note that since Q is a symmetric matrix, its eigenvalues $\lambda_i, i = 1, \dots, n$ are real and can be arranged in decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq \dots$. In particular, we have $\lambda_1 = 1$. The value $\delta = 1 - \max\{|\lambda_2|, |\lambda_n|\}$ is called the spectral gap.

In what follows, let K be an arbitrary constant that is not the same everywhere and may change even from one line to the next (of course, not causing any inconsistency).

For two functions $f(n), g(n)$ $g = O(f)$, if $\exists C$, a constant such that $|\frac{g}{f}| \leq C$, for large n , and $g = o(f)$ if $\limsup_{n \rightarrow \infty} |\frac{g}{f}| = 0$. Additionally, by $f \gg g$, we mean that $f > Cg$ for any constant C for n large enough.

An event E is said to hold with high probability (w.h.p.) if $\Pr(E) \geq 1 - O(n^{-c})$ for some $c > 0$. Recall that if a finite number of events hold true w.h.p., then so does their intersection. Furthermore, we say that a sequence of random variables in (Ω, \mathcal{F}, P) $X_n = o(1)$ w.h.p. if there exists a function $\psi(n) = o(1)$ such that the event $\{X_n \leq \psi(n)\}$ holds w.h.p.

In the present work, we consider families of random graphs with the following two properties:

Property I: W.h.p., $d_{max}^{(n)}/d_{min}^{(n)} \leq K$, where $d_{max}^{(n)}$ and $d_{min}^{(n)}$ are the maximum and minimum degrees, respectively.

Property II: W.h.p., $\max\{|\lambda_2^{(n)}|, |\lambda_n^{(n)}|\} = o(1)$.

The above two properties can be regarded as a variation of the expansion property. In the standard case of an expander family, one requires the graphs to be regular and the spectral gap $\delta = 1 - \max\{|\lambda_2|, |\lambda_n|\}$ to be bounded away from zero (see, e.g., [28]). Property I is a relaxation of the regularity condition, whereas Property II is stronger than the requirement for the spectral gap to be bounded away from zero. Properties I and II allow us to consider several standard families of random graphs such as Erdős-Rényi graphs, regular random graphs with increasing average degrees, and Chung-Lu graphs. For Chung-Lu graphs Property I imposes some restriction on the degree spread in the graph. It is worth noting that as a consequence of Property I we consider graphs that are always connected ($d_{min} > 0$) w.h.p.

Recall that the Personalized PageRank vector with a restart distribution vector v is defined as a stationary distribution of the modified Markov chain with the transition matrix

$$\tilde{P} = \alpha P + (1 - \alpha)v\mathbf{1}^T,$$

where α is a so-called damping factor [20]. We also recall the following useful formula for the Personalized PageRank π when $\alpha < 1$ (see, e.g., [22])

$$\pi = (1 - \alpha)[I - \alpha P]^{-1}v = (1 - \alpha)Rv. \quad (1)$$

3 Convergence in total variation

We recall that for two discrete probability distributions u and v , the total variation distance $d_{TV}(u, v)$ is defined as $d_{TV}(u, v) = \frac{1}{2} \sum_i |u_i - v_i|$. This can also be thought of as a 1-norm distance measure in the space of probability vectors, wherein for $x \in \mathbf{R}^n$, 1-norm $\|x\|_1 = \sum_i |x_i|$, and since for any probability vector π^n , $\|\pi^n\|_1 = 1 \forall n$, it makes sense to talk about convergence in 1-norm or TV-distance. Now we are in a position to formulate the following result.

PROPOSITION 1 *Let a family of graphs $G^{(n)}$ satisfy Properties I and II. If, in addition, $\|v\|_2 = O(1/\sqrt{n})$ ¹, the PageRank can be asymptotically approximated in total variation norm by a mixture of the restart distribution v and the vertex degree distribution. Namely, w.h.p.*

$$d_{TV}(\pi^{(n)}, \bar{\pi}^{(n)}) = o(1) \text{ as } n \rightarrow \infty,$$

where

$$\bar{\pi}^{(n)} = \frac{\alpha d^{(n)}}{\text{vol}(G^{(n)})} + (1 - \alpha)v,$$

and $\text{vol}(G^{(n)}) = \sum_i d_i^{(n)}$.

The above expression for the asymptotic PageRank vector is interesting in two respects: first it tells us that the rank vector asymptotically behaves like a convex combination of the stationary vector of a standard random walk with transition matrix P ; with the weight being α , and secondly, it starts to resemble the stationary vector as α tends to 1.

¹For a vector $x \in \mathbf{R}^n$, $\|x\|_2 = \sqrt{\sum_i |x_i|^2}$ is the 2-norm.

Proof: We only consider the case in which $0 < \alpha < 1$, since when $\alpha = 0$ or $\alpha = 1$, the statement of the theorem holds trivially.

We first note that the matrix Q can be written as follows by Spectral Decomposition Theorem [6]:

$$Q = u_1 u_1^T + \sum_{i=2}^n \lambda_i u_i u_i^T, \quad (2)$$

where $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues and $\{u_1, u_2, \dots, u_n\}$ are the corresponding orthogonal eigenvectors, $u_i \in \mathbf{R}^n$, $\|u_i\|_2 = 1$, and $u_1 = D^{1/2} \mathbf{1} / \sqrt{\mathbf{1}^T D \mathbf{1}}$ is the Perron–Frobenius eigenvector. Next, we rewrite (1) in terms of the matrix Q

$$\pi = (1 - \alpha) D^{1/2} [I - \alpha Q]^{-1} D^{-1/2} v.$$

Substituting (2) into the above equation, we obtain

$$\begin{aligned} \pi &= (1 - \alpha) D^{1/2} \left(\frac{1}{1 - \alpha} u_1 u_1^T + \sum_{i=2}^n \frac{1}{1 - \alpha \lambda_i} u_i u_i^T \right) D^{-1/2} v \\ &= D^{1/2} u_1 u_1^T D^{-1/2} v + (1 - \alpha) D^{1/2} \left(\sum_{i \neq 1} \frac{1}{1 - \alpha \lambda_i} u_i u_i^T \right) D^{-1/2} v. \end{aligned}$$

Let us denote the error by $\epsilon = \pi - \bar{\pi}$. Then, we can write

$$\begin{aligned} \epsilon &= \pi - \alpha D^{1/2} u_1 u_1^T D^{-1/2} v - (1 - \alpha) D^{1/2} I D^{-1/2} v \\ &= (1 - \alpha) D^{1/2} \left(\sum_{i \neq 1} \frac{u_i u_i^T}{1 - \alpha \lambda_i} - (I - u_1 u_1^T) \right) D^{-1/2} v \\ &= (1 - \alpha) D^{1/2} \left(\sum_{i \neq 1} u_i u_i^T \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} \right) D^{-1/2} v. \end{aligned}$$

Now let us bound the 1-norm $\|\epsilon\|_1$ of the error:

$$\begin{aligned} \|\epsilon\|_1 / (1 - \alpha) &= \left\| D^{1/2} \left(\sum_{i \neq 1} u_i u_i^T \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} \right) D^{-1/2} v \right\|_1 \\ &\leq C \sqrt{d_{\max} / d_{\min}} \sqrt{n} \max(|\lambda_2|, |\lambda_n|) \|v\|_2 \end{aligned} \quad (3)$$

by using $\frac{\|Ax\|_1}{\|x\|_2} \leq \sqrt{n} \frac{\|Ax\|_2}{\|x\|_2} \leq \sqrt{n} \|A\|_2$,² for any A and x , the submultiplicative property of matrix norm³ and the fact that $\|A\|_2 = \max_i |\lambda_i|$ if A is Hermitian [6]. Hence we have that $\|\epsilon\|_1 = o(1)$ w.h.p., under Properties I and II, when $\|v\|_2 = O(1/\sqrt{n})$. \square

Note that in the case of the standard PageRank, $v_i = 1/n$ implies $\|v\|_2 = O(1/\sqrt{n})$, but Theorem 1 also admits more general restart distributions than the uniform one.

²For any matrix $A \in \mathbf{R}^{m,n}$, $\|A\|_2 = \sup_{x, \|x\|_2=1} \|Ax\|_2$. [6]

³For two matrices $A \in \mathbf{R}^{m,n}$, and $B \in \mathbf{R}^{n,p}$, $\|AB\|_2 \leq \|A\|_2 \|B\|_2$,

COROLLARY 1 *The statement of Theorem 1 also holds with respect to the weak convergence, i.e., for any function f on V such that $\max_{x \in V} |f(x)| \leq 1$,*

$$\sup \left\{ \sum_v f(v) \pi_v - \sum_v f(v) \bar{\pi}_v \right\} = o(1) \quad w.h.p.$$

Proof: This follows from Theorem 1 and the fact that the left-hand side of the above equation is upper bounded by $2 d_{TV}(\pi_n, \bar{\pi}_n)$ [24]. ■

4 Chung-Lu random graphs

In this section, we analyze the asymptotics of the PageRank vector in random graphs. As a model of random graphs we consider the Chung-Lu model [13], which is a generalization of the Erdős–Rényi graph model, and hence our results naturally hold for Erdős–Rényi graphs also. The spectral properties of Chung-Lu graphs have been studied extensively in a series of papers [14, 15].

4.1 Chung-Lu random graph model

Let us first provide a definition of the Chung-Lu random graph model.

DEFINITION 1 (CHUNG-LU RANDOM GRAPH MODEL) *A Chung-Lu graph $\mathcal{G}(w)$ with an expected degree vector $w = (w_1, w_2, \dots, w_n)$, where w_i are positive real numbers, is generated by drawing an edge between any two vertices v_i and v_j independently of all others with probability $p_{ij} = \frac{w_i w_j}{\sum_k w_k}$, with the condition of existence $\max_i w_i^2 \leq \sum_k w_k$.*

Below we specify a corollary of Theorem 1 as applied to these graphs. But before that we need the following lemmas about Chung-Lu graphs mainly taken from [14, 15].

LEMMA 1 *If the expected degrees w_1, w_2, \dots, w_n satisfy $w_{\min} \gg \log(n)$, then in $\mathcal{G}(w)$ we have, w.h.p., $\max_i \left| \frac{d_i}{w_i} - 1 \right| = o(1)$.*

This result is shown in the sense of convergence in probability in [15], but it can be extended to the above result using Bernstein Concentration Lemma [7]:

$$\Pr\{|Y_n - \mathbb{E}Y_n| \geq \epsilon\} \leq 2 \exp \frac{-\epsilon^2}{2(B_n^2 + b\epsilon/3)},$$

where $B_n^2 = \mathbb{E}(Y_n - \mathbb{E}Y_n)^2$, $S_n = X_1 + \dots + X_n$, X_i are independent, and $|X_i| \leq b$. Applying this lemma to the degrees d_i of the Chung-Lu graph we see that

$$\Pr \left(\max_{1 \leq i \leq n} \left| \frac{d_i}{w_i} - 1 \right| \geq \beta \right) \leq \frac{2}{n^{c/4-1}}, \quad \text{if } \beta \geq \sqrt{\frac{c \log(n)}{w_{\min}}} = o(1)$$

if $w_{\min} \gg \log(n)$.

LEMMA 2 *If $w_{\max} \leq K w_{\min}$, and $\bar{w} = \sum_k w_k / n \gg \log^6(n)$, then for $\mathcal{G}(w)$ we have almost surely*

$$\|C\|_2 = \frac{2}{\sqrt{\bar{w}}} (1 + o(1)),$$

where $C = W^{-1/2} A W^{-1/2} - \chi^T \chi$, $W = \text{diag}(w)$, and $\chi_i = \sqrt{w_i / \sum_k w_k}$ is a row vector.

This lemma is an application of Theorem 5 in [14]. It can be verified that when $w_{max} \leq Kw_{min}$ and $\bar{w} \gg \log^6(n)$, the condition in Theorem 5 is satisfied, namely, $w_{min} \gg \sqrt{\bar{w}} \log^3(n)$, and hence the result follows.

LEMMA 3 For $\mathcal{G}(w)$ with $w_{max} \leq Kw_{min}$, and $\bar{w} \gg \log^6(n)$,

$$\max(\lambda_2(P), -\lambda_n(P)) = o(1) \quad \text{w.h.p.},$$

where P is the Markov matrix.

Proof: We have $\|Q - W^{-1/2}AW^{-1/2}\| = o(1)$ w.h.p. using Lemma 1 and the same argument as in the last part of the proof of Theorem 2 in [15]. From this, using Bauer-Fike Lemma [6], we get $|\lambda_i(P) - \lambda_i(W^{-1/2}AW^{-1/2})| = o(1)$ w.h.p. Then, using Lemma 3, we conclude that $\max_{i=1,2,\dots,n} |\lambda_i(C)| = o(1)$ for $\bar{w} \gg O(\log^6(n))$, as a consequence of Bauer-Fike Lemma and the fact that $\chi^T \chi$ is unit rank. So, $\max_{i=1,2,\dots,n} |\lambda_i(P)| \leq \max_{i=1,2,\dots,n} |\lambda_i(C)| + |\lambda_i(P) - \lambda_i(C)| = o(1)$ w.h.p. ■

COROLLARY 2 Let $\|v\|_2 = O(1/\sqrt{n})$, and $\alpha < 1$. Then PageRank π of the Chung-Lu graph can asymptotically be approximated in TV distance by $\bar{\pi}$, if $\bar{w} \gg \log^6(n)$ and $w_{max} \leq Kw_{min}$ for some K that does not depend on n .

Proof: Using Lemma 1 and the condition that $w_{max} \leq Kw_{min}$, one can show that $\exists K'$ s.t. $\frac{d_{max}}{d_{min}} \leq K'$ w.h.p. Then the result is a direct consequence of Lemma 3 and the inequality from (3). ■

We further note that this result holds also for Erdős-Rényi graphs $\mathcal{G}(n, p)$ with $np_n \gg \log^6(n)$, where we have $(w_1, w_2, \dots, w_n) = (np_n, np_n, \dots, np_n)$

4.2 Element-wise convergence

Earlier in this section, we proved the convergence of PageRank in TV distance for Chung-Lu random graphs. Note that since each component of PageRank could decay to zero as the graph size grows to infinity, this does not necessarily guarantee a convergence in an element-wise sense. In this section, we provide a proof for our convergence conjecture to include the element-wise convergence of the PageRank vector. Here we deviate slightly from the spectral decomposition technique and eigenvalue bounds used hitherto, and instead rely on well-known concentration bounds to bound the error in convergence.

Let $\bar{\Pi} = \text{diag}\{\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_n\}$ be a diagonal matrix whose diagonal elements are made of the components of the approximated PageRank vector and $\tilde{\delta} = \bar{\Pi}^{-1}(\pi - \bar{\pi})$, i.e., $\tilde{\delta}_i = (\pi_i - \bar{\pi}_i)/\bar{\pi}_i$. Then

$$\tilde{\delta}_i = \left((1 - \alpha)v_i + \alpha \frac{d_i}{\text{vol}(\mathcal{G})} \right)^{-1} \left[D^{1/2} \sum_{j \neq i} \frac{\alpha \lambda_j}{1 - \alpha \lambda_j} u_j u_j^T D^{-1/2} v \right]_i.$$

Therefore,

$$\|\tilde{\delta}\|_\infty \leq \frac{\sum_i d_i/n}{\alpha d_{\min}} \sqrt{d_{\max}} \left\| \sum_{i \neq 1} \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} u_i u_i^T \tilde{v}' \right\|_\infty, \quad (4)$$

where $\tilde{v}' \equiv nD^{-1/2}v$, and $\|\tilde{\delta}\|_\infty = \max_i |\tilde{\delta}_i|$.

Define $\tilde{Q} = Q - u_1 u_1^T$, the restriction of the matrix Q to the orthogonal subspace of u_1 . Later in this section we prove the following lemma.

LEMMA 4 For a Chung-Lu random graph $\mathcal{G}(w)$ with expected degrees w_1, w_2, \dots, w_n , where $w_{\max} \leq Kw_{\min}$ and $w_{\min} \gg \log(n)$, we have with high probability, $\|\tilde{Q}\tilde{v}'\|_\infty = o(1/\sqrt{w_{\min}})$, when $v_i = O(1/n) \forall i$.

The next lemma is related to the matrix $S = (I - \alpha Q)^{-1}$, as defined earlier in the paper.

LEMMA 5 Under the conditions of Lemma 4, $\|S\|_\infty \leq C$ w.h.p., where C is a number independent of n that depends only on α and K .

Proof: Note that $S = (I - \alpha Q)^{-1} = D^{-1/2}(I - \alpha P)^{-1}D^{1/2}$. Therefore, $\|S\|_\infty \leq \sqrt{\frac{d_{\max}}{d_{\min}}} \|(I - \alpha P)^{-1}\|_\infty$ and the result follows since $\|(I - \alpha P)^{-1}\|_\infty \leq \frac{1}{1-\alpha}$ [22] and using Lemma 1. ■

PROPOSITION 2 Let $v_i = O(1/n) \forall i$, and $\alpha < 1$. PageRank π converges element-wise to $\bar{\pi} = (1 - \alpha)v + \alpha d / \text{vol}(G)$, in the sense that $\max_i (\pi_i - \bar{\pi}_i) / \bar{\pi}_i = o(1)$ w.h.p., on the Chung-Lu graph with expected degrees $\{w_1, w_2, \dots, w_n\}$ such that $w_{\min} > \log^c(n)$ for some $c > 1$ and $w_{\max} \leq Kw_{\min}$, K being a constant independent of n .

Proof: Define $Z := \sum_{i \neq 1} \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} u_i u_i^T$. We then have:

$$\begin{aligned} Z &= \sum_{i=1}^n \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} u_i u_i^T - \frac{\alpha}{1 - \alpha} u_1 u_1^T = (I - \alpha Q)^{-1} \alpha Q - \frac{\alpha}{1 - \alpha} u_1 u_1^T \\ &= S \left[\alpha Q - \frac{\alpha}{1 - \alpha} (I - \alpha Q) u_1 u_1^T \right] = \alpha S \tilde{Q} \end{aligned}$$

Using Lemmas 5 and 4, and equation (4), we have:

$$\|\tilde{\delta}\|_\infty = C \frac{\sum_i d_i / n}{d_{\min}} \sqrt{d_{\max}} o(1/\sqrt{w_{\min}}) = C \left(\frac{d_{\max}}{d_{\min}} \right)^{\frac{3}{2}} o(1), \quad (5)$$

which using Lemma 1 is $o(1)$ w.h.p. ■

COROLLARY 1 (ERDŐS-RÉNYI GRAPHS) For an Erdős-Rényi graph $G(n, p)$ with $np_n \gg \log(n)$, we have that asymptotically the personalized PageRank π converges pointwise to $\bar{\pi}$ for v such that $v_i = O(1/n)$.

Proof for Lemma 4: From Lemma 1, we have for Chung-Lu graphs that: $d_i = w_i(1 + \epsilon_i)$, where $\omega \equiv \max_i \epsilon_i = o(1)$ with high probability. In the proof we assume explicitly that $v_i = 1/n$, but the results hold in the slightly more general case where $v_i = O(1/n)$ uniformly $\forall i$, i.e., $\exists K$ such that $\max_i n v_i \leq K$. It can be verified easily that all the bounds that follow hold in this more general setting. The event $\{\omega = o(1)\}$, holds w.h.p. asymptotically from Lemma 1. In this case, we have

$$\sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_i d_i} \right) \frac{v_j}{\sqrt{d_j}} = \sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_k d_k} \right) \frac{v_j}{\sqrt{w_j}} (1 + \epsilon_j)$$

where ϵ_j is the error of convergence, and we have $\max_j \epsilon_j = O(\omega)$. Therefore,

$$\|\tilde{Q}\tilde{v}\|_\infty \leq \|\tilde{Q}q\|_\infty + \max_i \epsilon_i \|\tilde{Q}q\|_\infty \leq \|\tilde{Q}q\|_\infty (1 + o(1)) \quad \text{w.h.p.,}$$

where q is a vector such that $q_i = \frac{nw_i}{\sqrt{w_i}}$. Furthermore, we have w.h.p.

$$\begin{aligned} \frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_k d_k} &= \frac{A_{ij}}{\sqrt{w_i(1+\epsilon_i)w_j(1+\epsilon_j)}} - \frac{\sqrt{w_i(1+\epsilon_i)w_j(1+\epsilon_j)}}{\sum_k w_k(1+\epsilon_k)} \\ &= \frac{A_{ij}}{\sqrt{w_i w_j}} (1 + O(\epsilon_i) + O(\epsilon_j)) - \frac{\sqrt{w_i w_j}}{\sum_k w_k} \left(\frac{1 + O(\epsilon_i) + O(\epsilon_j)}{1 + O(\omega)} \right) \\ &= \left(\frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k} \right) (1 + \delta_{ij}), \end{aligned}$$

where δ_{ij} is the error in the ij^{th} term of the matrix and $\delta_{ij} = O(\omega)$ uniformly, so that $\max_{ij} \delta_{ij} = o(1)$ w.h.p. Consequently, defining $\tilde{Q}_{ij} = \frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}$ we have:

$$\begin{aligned} \|\tilde{Q}q\|_\infty &\leq \|\tilde{Q}q\|_\infty + \max_i \left| \sum_j \tilde{Q}_{ij} \delta_{ij} q_j \right| \\ &\leq \|\tilde{Q}q\|_\infty + O(\omega) \max_i \frac{1}{\sqrt{w_{\min}}} \sum_j |\tilde{Q}_{ij}| \\ &\leq \|\tilde{Q}q\|_\infty + o(1) \frac{1}{\sqrt{w_{\min}}} \left(\sqrt{\frac{w_{\max}}{w_{\min}}} + \frac{w_{\max}}{w_{\min}} \right) \end{aligned} \quad (6)$$

$$\leq \|\tilde{Q}q\|_\infty + o(1/\sqrt{w_{\min}}) \quad (7)$$

where in (6) we used the fact the $O(\omega)$ is a uniform bound on the error and it is $o(1)$, and the fact that $\max_i \sum_j |\tilde{Q}_{ij}| \leq \max_i \sum_j \frac{A_{ij}}{\sqrt{w_i w_j}} + \sum_j \frac{\sqrt{w_i w_j}}{\sum_k w_k} \leq \sqrt{\frac{w_{\max}}{w_{\min}}} + \frac{w_{\max}}{w_{\min}}$ from simple bounds, and $\max_j q_j \leq \frac{1}{\sqrt{w_{\min}}}$. Now we proceed to bound $\|\tilde{Q}q\|_\infty$. Substituting for $q_i = \frac{1}{\sqrt{w_i}}$, we get

$$\sum_j \frac{1}{\sqrt{w_j}} \left(\frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k} \right) = \sum_j \frac{1}{w_j \sqrt{w_i}} \left(A_{ij} - \frac{w_i w_j}{\sum_i w_i} \right) \equiv \frac{1}{\sqrt{w_i}} X_i. \quad (8)$$

We seek to bound $\max_i |X_i|$: $X_i = \sum_j \frac{1}{w_j} \left(A_{ij} - \frac{w_i w_j}{\sum_i w_i} \right) \equiv Y_n - \mathbb{E}Y_n$, where $Y_n = \sum_j \frac{A_{ij}}{w_j}$. Furthermore, $\mathbb{E}X_n^2 = \sum_j \frac{1}{w_j^2} \mathbb{E}(A_{ij} - p_i)^2$, with $p_i = \frac{w_i w_j}{\sum_i w_i}$. So, $\mathbb{E}S_n^2 \leq \frac{w_i}{\sum_i w_i} \sum_j \frac{1}{w_j} \leq n \frac{p_i}{w_{\min}}$, and $\frac{A_{ij}}{w_j} \leq 1/w_{\min}$. Therefore by use of Bernstein Concentration Lemma for $\epsilon < n \max_i p_i$:

$$\begin{aligned} \Pr \left\{ \max_i \left| \sum_j (A_{ij} - p_{ij})/w_j \right| \geq \epsilon \right\} &\leq n \max_i \exp \left(-\frac{\epsilon^2}{2(p_i n/w_{\min}) + \epsilon/w_{\min}} \right) \\ &\leq n \max_i \exp \left(-\frac{w_{\min} \epsilon^2}{2(np_i + \epsilon)} \right) \leq n \exp \left(-\epsilon^2 w_{\min} / (4n \max_i p_i) \right) \\ &\leq n \exp \left(\frac{-\epsilon^2 \text{vol} w_{\min}}{4w_{\max} n} \right), \end{aligned}$$

where $\frac{\text{vol}}{n} = \frac{\sum_i w_i}{n} \geq w_{\min}$. It can be verified that when $\epsilon = \frac{1}{(\bar{w})^\alpha}$ for example, for some $\alpha > 0$, the probability above can be upper bounded by $n^{-(\gamma K-1)}$, if $\bar{w} \geq (\gamma \log(n))^{\frac{1}{1-2\alpha}}$, for some γ large enough, which can be easily satisfied if $w_{\min} \gg O(\log^c(n))$, for some $c > 1$. Thus,

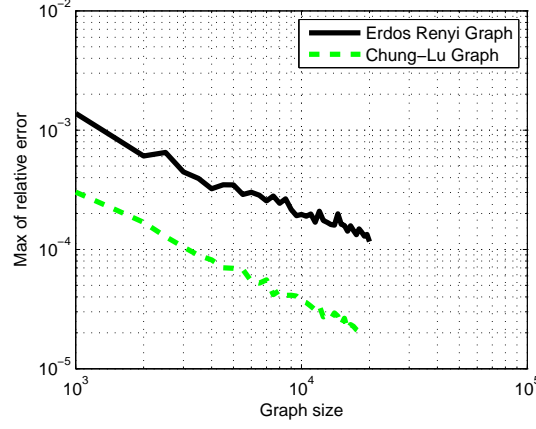


Figure 1: Log-log plot of maximum relative error for Erdős-Rényi and Chung-Lu graphs

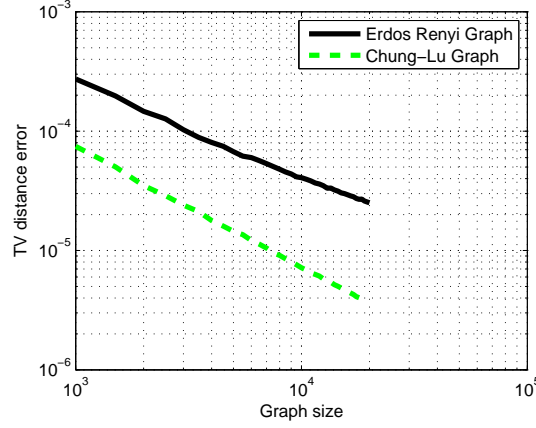


Figure 2: Log-log plot of TV distance error for Erdős-Rényi and Chung-Lu graphs

finally, from (8) and (7) we have $\|\tilde{Q}q\|_{\infty} = o(1/\sqrt{w_{\min}})$, w.h.p., and therefore from (6), we get $\|\tilde{Q}\tilde{v}'\|_{\infty} = o(1/\sqrt{w_{\min}})$. ■

5 Experimental results

In this section, we provide experimental evidence to further illustrate the analytic results obtained in the previous sections. In particular, we simulated Erdős-Rényi graphs with $p_n = C \frac{\log^7(n)}{n}$ and Chung-Lu graphs with the degree vector w sampled from a geometric distribution so that the average degree $\bar{w} = O(n^{1/3})$, clipped such that $w_{\max} = 7w_{\min}$, for various values of graph size, and plotted the maximum of relative error $\tilde{\delta}$ and TV distance error $\|\delta\|_1$, respectively, in Figures 1 and 2. As expected, both these errors decay as functions of the graph size, which illustrates that the PageRank vector does converge to the asymptotic value. In the interest of

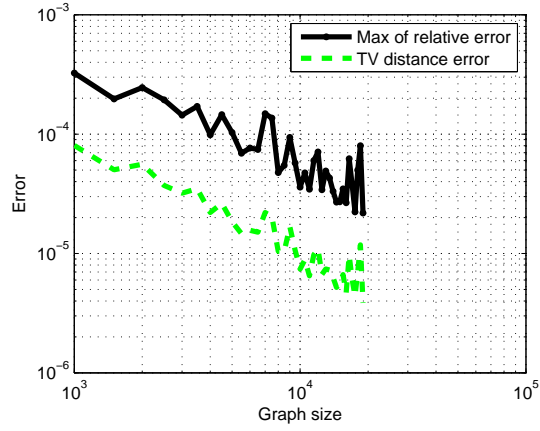
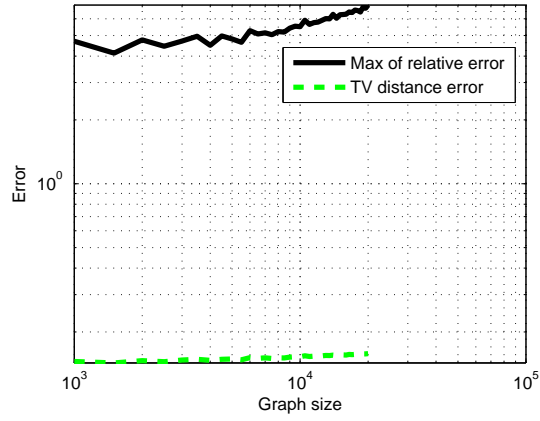


Figure 3: Log-log plot of TV distance and maximum relative error for power-law graphs

Figure 4: Log-log plot of TV distance and maximum relative error for ER-graph when $v = e_1$

exploration, we also conducted simulations on power-law graphs with exponent 5, and it appears that PageRank converges for these graphs as well, albeit the decay of the error being slightly noisier than observed in the previous examples (see Figure 3). This requires further study.

Furthermore, we see that when $v_i = 1$ for some i the convergence doesn't hold (Figure 4 in the case of Erdős–Rényi graphs). Whereas we see from our analysis that if $v_k = 1$ for some k , the quantity $\left\| \tilde{Q}D^{-1/2}v \right\|_{\infty}$, becomes:

$$\max_i \left| \sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_l d_l} \right) v_j / \sqrt{d_j} \right| = \max_i \frac{1}{\sqrt{d_i d_k}} \left| A_{ik} - \frac{d_i d_k}{\sum_l d_l} \right|,$$

which is $O\left(\frac{1}{\sqrt{w_{\min} w_k}}\right)$ and does not fall sufficiently fast.

6 Conclusions

In this work, we showed that when the size of a graph tends to infinity, the PageRank vector lends itself to be approximated by a mixture of the restart distribution and the degree distribution, subject to some conditions for a class of random graphs. In future, we would like to relax some of these conditions, especially the condition on degree spread. This condition can be relaxed as demonstrated by simulations on power-law graphs. In addition, we would like to obtain closer bounds on the error for Chung-Lu graphs as empirical evidence suggests these can be improved too.

Acknowledgements

We would like to thank Nelly Litvak for stimulating discussions on the topic of the paper. This work was partially supported by the French Government (National Research Agency, ANR) through the “Investments for the Future” Program reference ANR-11-LABX-0031-01 and ADR “Network Science” from Joint Inria Alcatel-Lucent Lab.

References

- [1] Andersen, R., Chung, F., and Lang, K. Local graph partitioning using pagerank vectors. In *Proceedings of IEEE FOCS*, 2006.
- [2] Avrachenkov, K., Cottatellucci, L., and Kadavankandy, A. Spectral properties of random matrices for stochastic block model. In *Proceedings of WiOpt Workshop PhysComNet*, 2015.
- [3] Avrachenkov, K., Dobrynin, V., Nemirovsky, D., Pham, S.K., and Smirnova, E. Pagerank based clustering of hypertext document collections. In *Proceedings of ACM SIGIR*, pp. 873–874, 2008.
- [4] Avrachenkov, K., Gonçalves, P., Mishenin, A., and Sokol, M. Generalized optimization framework for graph-based semi-supervised learning. In *Proceedings of SIAM Conference on Data Mining*, vol. 9, 2012.
- [5] Avrachenkov, K., and Lebedev, D. PageRank of scale-free growing networks. *Internet Mathematics*, 3(2), 207-231, 2006.
- [6] Bhatia, R. *Matrix analysis*. Springer Science and Business Media, vol. 169, 2013.
- [7] Billingsley, P. *Probability and measure*. John Wiley & Sons, 2008.
- [8] Boudin, F. A comparison of centrality measures for graph-based keyphrase extraction. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, 2013.
- [9] Chen, N., Litvak, N., and Olvera-Cravioto, M. PageRank in scale-free random graphs. In *Proceedings of WAW*, 2014.
- [10] Chen, N., Litvak, N., and Olvera-Cravioto, M. Ranking algorithms on directed configuration networks. Preprint, arXiv:1409.7443, 2014.
- [11] Chung, F. A local graph partitioning algorithm using heat kernel PageRank. In *Proceedings of WAW* 2009.

- [12] Chung, F. Spectral graph theory. American Mathematical Soc, v.92, 1997.
- [13] Chung, F., and Lu, L. The average distances in random graphs with given expected degrees. PNAS, v.99(25), 15879-15882, 2002.
- [14] Chung, F., Lu, L., and Vu, V. Spectra of random graphs with given expected degrees. PNAS, v.100(11), 6313-6318, 2003.
- [15] Chung, F., and Radcliffe, M. On the spectra of general random graphs. The electronic journal of combinatorics, 18(1), 2011.
- [16] Ding, C., He, X., Husbands, P., Zha, H., and Simon, H.D. PageRank, HITS and a unified framework for link analysis. In Proceedings of ACM SIGIR, 2002.
- [17] Erdős, P., and Rényi, A. On random graphs. Publicationes Mathematicae Debrecen, v.6, 290-297, 1959.
- [18] Fortunato, S., Boguna, M., Flammini, A., and Menczer, F. Approximating PageRank from in-degree. In Proceedings of WAW, 2008.
- [19] Gkorou, D., Vinko, T., Pouwelse, J., and Epema, D. Leveraging node properties in random walks for robust reputations in decentralized networks. In Proceedings of IEEE Peer-to-Peer Computing (P2P), 2013.
- [20] Haveliwala, T.H. Topic-sensitive pagerank. In Proceedings of WWW, 517-526, 2002.
- [21] Kamvar, S.D., Schlosser, M.T., and H. Garcia-Molina, H. The eigentrust algorithm for reputation management in p2p networks. In Proceedings of WWW, 2003.
- [22] Langville, A.N., and Meyer, C.D. Deeper inside pagerank. Internet Mathematics, 1(3), 335-380, 2004.
- [23] Litvak, N., Scheinhardt, W. R., and Volkovich, Y. In-degree and PageRank: Why do they follow similar power laws? *Internet mathematics*, 4(2-3), 175-198, 2007.
- [24] Levin, D.A., Peres, Y., and Wilmer, E.L. Markov chains and mixing times. American Mathematical Soc., 2009.
- [25] Moler, C., Moler, K. Numerical Computing with MATLAB. SIAM, 2003.
- [26] Page, L., Brin, S., Motwani, R., and Winograd, T. PageRank: Bringing order to the web. Stanford Digital Libraries Working Paper, v.72, 1997.
- [27] Perra, N. and Fortunato, S. Spectral centrality measures in complex networks. Phys. Rev. E 78, 036107, 2008.
- [28] Vadhan, S. Pseudorandomness. Foundations and Trends in Theoretical Computer Science, v.7(1-3), pp.1-336, Now publishers, 2012.
- [29] Volkovich, Y., and Litvak, N. Asymptotic analysis for personalized web search. *Adv. Appl. Prob.*, 42(2), 577-604, 2010.
- [30] Yeh, E., Ramage, D., Manning, C. D., Agirre, E., and Soroa, A. WikiWalk: random walks on Wikipedia for semantic relatedness. In Proceedings of the Workshop on Graph-based Methods for Natural Language Processing, 2009.

Contents

1	Introduction	3
2	Definitions	3
3	Convergence in total variation	4
4	Chung-Lu random graphs	6
4.1	Chung-Lu random graph model	6
4.2	Element-wise convergence	7
5	Experimental results	10
6	Conclusions	12



**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

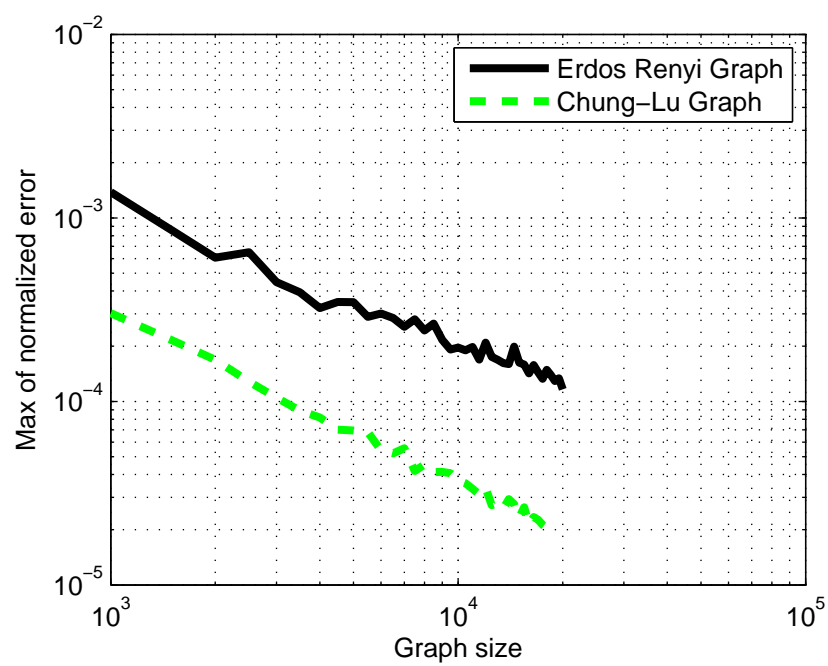
ISSN 0249-6399

This figure "logo-inria.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1511.04925v1>

This figure "logo-inria.png" is available in "png" format from:

<http://arxiv.org/ps/1511.04925v1>



This figure "pagei.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1511.04925v1>

This figure "pagei.png" is available in "png" format from:

<http://arxiv.org/ps/1511.04925v1>

This figure "rrpage1.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1511.04925v1>

This figure "rrpage1.png" is available in "png" format from:

<http://arxiv.org/ps/1511.04925v1>